Wireless Device-to-Device Communications with Distributed Caching

Negin Golrezaei, Alexandros G. Dimakis, Andreas F. Molisch
Dept. of Electrical Eng.
University of Southern California
Los Angeles, CA, USA
emails: {golrezae,dimakis,molisch}@usc.edu

***Abstract*—We introduce a novel wireless device-to-device (D2D) collaboration architecture that exploits distributed storage of popular content to enable frequency reuse. We identify a fundamental conflict between collaboration distance and interference and show how to optimize the transmission power to maximize frequency reuse. Our analysis depends on the user content request statistics which are modeled by a Zipf distribution. Our main result is a closed form expression of the optimal collaboration distance as a function of the content reuse distribution parameters. We show that if the Zipf exponent of the content reuse distribution is greater than 1, it is possible to have a number of D2D interference-free collaboration pairs that scales linearly in the number of nodes. If the Zipf exponent is smaller than 1, we identify the best possible scaling in the number of D2D collaborating links. Surprisingly, a very simple distributed caching policy achieves the optimal scaling behavior and therefore there is no need to centrally coordinate what each node is caching.**

I. Introduction

Wireless mobile data traffic is expected to increase by a factor of 40 over the next five years, from the current 93 Petabytes to 3600 Petabytes per month in the next five years [1]. This explosive demand is fueled mainly by mobile video traffic that is expected to increase by a factor of 65 times, and become the by far dominant source of data traffic. Modern smartphones and tablets have significant storage capacity often reaching several gigabytes. Recent breakthroughs in dense NAND flash will make 128GB smartphone memory chips available in the coming months. In this paper we show how to exploit these storage capabilities to significantly reduce wireless capacity bottlenecks.

The central idea in this paper is that, for most types of mobile video traffic, we can replace backhaul connectivity with storage capacity. This is true because of *content reuse*, *i.e.*, the fact that popular video files will be requested by a large number of users. Distributed storage enhances the opportunities for user collaboration.

We recently introduced the idea of femtocaching helpers [2] [3], small base stations with a low-bandwidth (possibly wireless) backhaul link and high storage capabilities. In this paper we take this architecture one step further: We introduce *a device-to-device (D2D) architecture where the mobiles are used as caching storage nodes*. Users can collaborate by caching popular content and utilizing local device-to-device communication when a user in the vicinity requests a popular file. The base station can keep track of the availability of the cached content and direct requests to the most suitable nearby device. Storage allows users to collaborate even when they do not request the same content *at the same time*. This is a new dimension in wireless collaboration architectures beyond relaying and cooperative communications.

This research was supported in part by NSF Career Grant CCF-1055099 and research gifts by Intel and Microsoft Research.

Our contributions: In this paper we introduce the novel D2D architecture and formulate some theoretical problems that arise. Specifically, we identify a conflict between collaboration distance and interference. We show how to optimize the D2D collaboration distance and analyze the scaling behavior of D2D benefits. The optimal collaboration distance depends on the content request statistics which are modeled by a Zipf distribution. Our main result is a closed form expression of the optimal collaboration distance as a function of the content reuse distribution parameters. We show that if the Zipf exponent of the content reuse distribution is greater than 1, it is possible to have a number of D2D interference-free collaboration pairs that scales linearly in the number of nodes.

If the Zipf exponent is smaller than 1, we identify the best possible scaling in the number of D2D collaborating links. Surprisingly, a very simple distributed caching policy achieves the optimal scaling behavior and therefore there is no need to centrally coordinate what each node is caching.

The remainder of this paper is organized as follows: In Section II we setup the D2D formulation and explain the tradeoff between collaboration distance and interference. Section III contains our two main theorems, the scaling behavior for Zipf exponents greater and smaller than 1. In Section IV we discuss future directions, open problems and conclusions. Finally, in the Appendix we include some interesting technical parts of our proofs. Due to space constraints we omit the complete proofs from this version of the paper.

II. Model and Setup

We consider n users distributed uniformly in a unit square and consider this as single cell. The base station (BS) might be aware of the stored files and channel state information of the

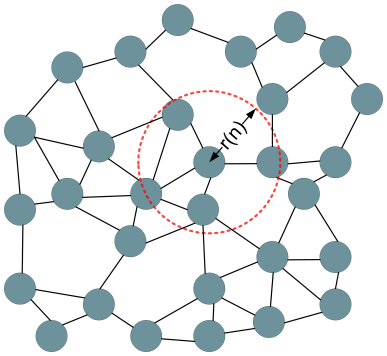


Fig. 1. Random geometric graph example with collaboration distance $r(n)$.

users and control the D2D communications. For simplicity, we neglect inter-cell interference and consider one cell in isolation. We further assume that the D2D communication does not interfere with communication between the BS and users. This assumption is justified if the D2D communications occur in a separate frequency band (*e.g.*, WiFi). For the device-to-device throughput, we henceforth do not need to consider explicitly the BS and its associated communications.

The communication is modeled by random geometric graph $G(n, r(n))$ where two users (assuming D2D communication is possible) can communicate if their physical distance is smaller than some collaboration distance $r(n)$ [4], [5]. The maximum allowable distance for D2D communication $r(n)$ is determined by the power level for each transmission. Figure 1 illustrates an example of random geometric graph (RGG).

We assume that users may request files from a set of size m that we call a "library". The size of this set should increase as a function of the number of users n. Intuitively, the set of YouTube videos requested in Berkeley in one day should be smaller than the set of requested in Los Angeles. We assume that this growth should be sublinear in n, *e.g.* m could be $\Theta(\log(n))$.

Each user requests a file from the library by sampling independently using a popularity distribution. Based on numerous studies, Zipf distributions have been established as good models to the measured popularity of video files [6], [7]. Under this model, the frequency of the ith popular file, denoted by f_i, is inversely proportional to its rank:

$$f_i = \frac{\frac{1}{i^{\gamma_r}}}{\sum_{j=1}^{m} \frac{1}{j^{\gamma_r}}}, \quad 1 \leq i \leq m. \tag{1}$$

The Zipf exponent γ_r characterizes the distribution by controlling the relative popularity of files. Larger γ_r exponents correspond to higher content reuse, *i.e.,* the first few popular files account for the majority of requests.

Each user has a storage capacity called cache which is populated with some video files. For our scaling law analysis we assume that all files have the same size, and each user can store one file. This yields a clean formulation and can be easily extended for larger storage capacities.

Our architecture works as follows: If a user requests one of the files stored in neighbors' caches in the RGG, neighbors will handle the request locally through D2D communication; otherwise, the BS should serve the request. Thus, to have D2D communication it is not sufficient that the distance between two users be less than $r(n)$; users should find their desired files locally in caches of their neighbors. A link between two users will be called potentially active if one requests a file that the other is caching. Therefore, the probability of D2D collaboration opportunities depends on what is stored and requested by the users.

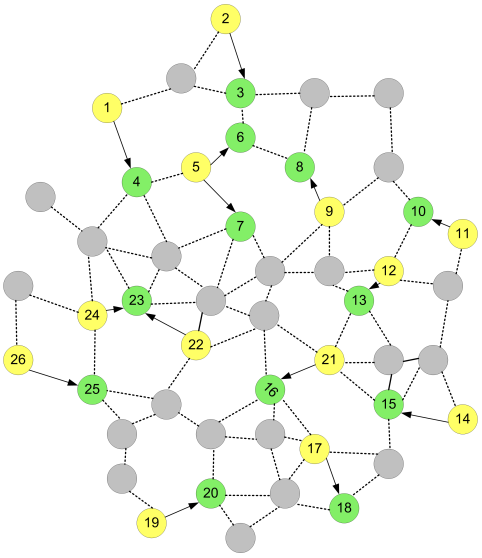
The decision of what to store can be taken in a distributed or centralized way. A central control of the caching by the BS allows very efficient file-assignment to the users [8]. However, if such control is not desired or the users are highly mobile, caching has to be optimized in a distributed way. The simple randomized caching policy we investigate makes each user choose which file to cache by sampling from a caching distribution. It is clear that popular files should be stored with a higher probability, but the question is that how much redundancy we want to have in our distributed cache.

We assume that all D2D links share the same time-frequency transmission resource within one cell area. This is possible since the distance between requesting user and user with the stored file will typically small. However, there should be no destructive interference of a transmission by others on an active D2D link. We assume that (given that node u wants to transmit to node v) any transmission within range $r(n)$ from v (the receiver) can introduce interference for the $u-v$ transmission. Thus, they cannot be activated simultaneously. This model is known as *protocol model*; while it neglects important wireless propagation effects such as fading [9], it can provide fundamental insights and has been widely used in prior literature [4].

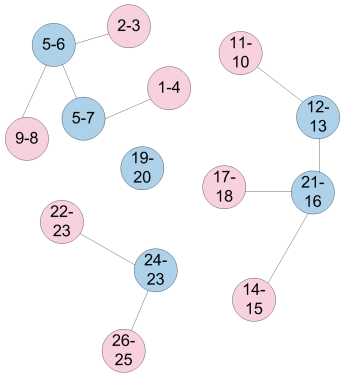
To model interference given a storage configuration and user requests we start with all potential D2D collaboration links. Then, we construct the conflict graph as follows. We model any possible D2D link between node u as transmitter to node v as a receiver with a vertex $u-v$ in the conflict graph. Then, we draw an edge between any two vertices (links) that create interference for each other according to the protocol model. Figure 2 shows how the RGG is converted to the conflict graph. In Figure 2(a), receiver nodes are green and transmitter nodes are yellow. The nodes that should receive their desired files from the BS are gray. A set of D2D links is called active if they are potentially active and can be scheduled simultaneously, *i.e.*, form an independent set in the conflict graph. The random variable counting the number of active D2D links under some policy is denoted by L.

Figure 2(b) shows the conflict graph and one of maximum independent sets for the conflict graph. We can see that out of 14 possible D2D links 9 links can co-exist without interference. As is well known, determining the maximum independent set of an arbitrary graph is computationally intractable (NP complete [10]). Despite the difficulty of characterizing the number of interference-free active links, we can determine

the best possible scaling law in our random ensemble.



(a)



(b)

Fig. 2. a) Random geometric graph, yellow and green nodes indicate receivers, transmitters in D2D links. Gray nodes get their request files from the BS. Arrows show all possible D2D links. b) conflict graph based on Figure 2(a) and one of maximum independent set of the conflict graph; pink vertices are those D2D links that can be activated simultaneously.

III. Analysis

A. Finding the optimal collaboration distance

We are interested in determining the best collaboration distance $r(n)$ and caching policy such that the expected number of active D2D links is maximized. Our optimization is based on balancing the following tension: The smaller the transmit power, the smaller the region in which a D2D communication creates interference. Therefore, more D2D pairs can be packed into the same area allowing higher frequency reuse. On the other hand, a small transmit power might not be sufficient to reach a mobile that stores the desired file. Smaller power means smaller distance and hence smaller probability of collaboration opportunities. The optimum way to solve this problem would be to assign different transmit power to each node dynamically, to maximize the number of non-interfering collaborating pairs. However this approach would be intractable and non-practical.

Our approach is to enforce the same transmit power for all the users and show how to optimize it based on the content request statistics. Our analysis involves finding the best compromise between the number of possible parallel D2D links and the probability of finding the requested content. Our results consist of two parts. In the first part (upper bound), we find the best achievable scaling for the expected number of active D2D links. In the second part (achievability), we determine an optimal caching policy and $r(n)$ to obtain the best scaling for the expected number of active links $E[L]$.

The best achievable scaling for the expected number of active D2D links depends on the extend of content reuse. Larger Zipf distribution exponents correspond to more redundancy in the user requests and a small number of files accounts for the majority of video traffic. Thus, the probability of finding requested files through D2D links increases by having access to few popular files via neighbors.

We separate the problem into two different regions depending on the Zipf exponent: $\gamma_r > 1$ and $\gamma_r < 1$. For each of these regimes, we find the best achievable scaling for $E[L]$ and the optimum asymptotic $r(n)$ denoted by $r_{opt}(n)$. We also show that a simple distributed caching policy with the properly chosen caching distribution has optimal scaling, *i.e.*, matches the scaling behavior that any centralized caching policy could achieve[1].

Our first result is the following theorem:

Theorem 1: If the Zipf exponent $\gamma_r > 1$,

i) **Upper bound:** For any caching policy, $E[L] = O(n)$,

ii) **Achievability:** Given that $\sqrt{\frac{c_1}{n}} \leq r_{opt}(n) \leq \sqrt{\frac{c_2}{n}}$ and using a Zipf caching distribution with exponent $\gamma_c > 1$ then $E[L] = \Theta(n)$.

The first part of the theorem 1 is trivial since the number of active D2D links can at most scale linearly in the number of users. The second part indicates that if we choose $r_{opt}(n) = \Theta(\sqrt{\frac{1}{n}})$ and $\gamma_c > 1$, $E[L]$ can grow linearly with n. There is some simple intuition behind this result: We show that in this regime users are surrounded by a constant number of users in expectation. If the Zipf exponent γ_c is greater than one, this suffices to show that the probability that they can find their desired files locally is a non-vanishing constant as n grows. Our proof is provided in the Appendix A.

For the low content reuse region $\gamma_r < 1$, we obtain the following result:

Theorem 2: If $\gamma_r < 1$,

i) **Upper bound:** For any caching policy, $E[L] = O(\frac{n}{m^\eta})$ where $\eta = \frac{1-\gamma_r}{2-\gamma_r}$,

ii) **Achievability:** If $r_{opt}(n) = \Theta(\sqrt{\frac{m^{\eta+\epsilon}}{n}})$ and users cache files randomly and independently according to a Zipf dis-

[1]We use the standard Landau notation: $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$ respectively denote $|f(n)| \leq c_1 g(n)$ and $|f(n)| \geq c_2 g(n)$ for some constants c_1, c_2. $f(n) = \Theta(g(n))$, stands for $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$. Little-o notation, *i.e.*, $f(n) = o(g(n))$ is equivalent to $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$.

tribution with exponent γ_c, for any exponent $\eta+\epsilon$, there exists γ_c such that $E[L]=\Theta(\frac{n}{m^{\eta+\epsilon}})$ where $0<\epsilon<\frac{1}{6}$ and γ_c is a solution to the following equation

$$\frac{(1-\gamma_r)\gamma_c}{1-\gamma_r+\gamma_c}=\eta+\epsilon.$$

We show that when there is low content reuse, linear scaling in frequency re-use is not possible. At a high level, in order to achieve the optimal scaling, on average a user should be surrounded by $\Theta(m^\eta)$ users. Comparing with the first region where $\gamma_r>1$, we can conclude that when there is less redundancy, users have to see more users in the neighborhood to find their desired files locally. Due to space constraints we omit this proof.

IV. Discussion and Conclusions

The study of scaling laws of the capacity of wireless networks has received significant attention since the pioneering work by Gupta and Kumar [4] (*e.g.* see [11]–[13]). The first result was pessimistic: if n nodes are trying to communicate (say by forming $n/2$ pairs), since the typical distance in a 2D random network will involve roughly $\Theta(\sqrt{n})$ hops, the throughput per node must vanish, approximately scaling as $1/\sqrt{n}$. There are, of course, sophisticated arguments performing rigorous analysis that sharpens the bounds and numerous interesting model extensions. One that is particularly relevant to this project is the work by Grossglauser and Tse [12] that showed that if the nodes have infinite storage capacity, full mobility and there is no concern about delay, constant (non-vanishing) throughput per node can be sustained as the network scales.

Despite the significant amount of work on ad hoc networks, there has been very little work on file sharing and content distribution over wireless ([2], [14]) beyond the multiple unicast traffic patters introduced in [4]. Our result shows that if there is sufficient content reuse, non-vanishing throughput per node can be achieved, even with constant storage and delay. In our recent work [15] we empirically analyzed the optimal collaboration distance for fixed number of users.

On a more technical note, the most surprising result is perhaps the fact that in Theorem 2, a simple distributed policy can match the optimal scaling behavior $E[L]=O(\frac{n}{m^\eta})$. Further, for both regimes, the distributed caching policy exponent γ_c should not match the request Zipf exponent γ_r, something that we found quite counter intuitive.

Overall, even if linear frequency re-use is not possible, we expect the scaling of the library m to be quite small (typically logarithmic) in the number of users n. In this case we obtain near-linear (up to logarithmic factors) growth in the number of D2D links for the full spectrum of Zipf exponents. Our results are encouraging and show that distributed caching can enable collaboration and mitigate wireless content delivery problems.

Appendix A
Proof of Theorem 1

The first part of the theorem is easy to see since the number of D2D links cannot exceed the number of users.

For the second part of theorem 1, we divide the cell into $\frac{2}{r(n)^2}$ virtual square clusters. Figure 3(a) shows the virtual clusters in the cell. The cell side is normalized to 1 and the side of each cluster is equal to $\frac{r(n)}{\sqrt{2}}$. Thus, all users within a cluster can communicate with each other. Based on our interference model, in each cluster only one link can be activated. Thus, to prove the theorem, it is enough to show that in a constant fraction of virtual clusters, there are active D2D links that do not introduce interference to each other. This is because $r(n)=\Theta(\sqrt{\frac{1}{n}})$ and there are $\Theta(n)$ virtual clusters in the cell. When there is an active D2D link within a cluster, we call the cluster *good*. But not all good clusters can be activated simultaneously. One good cluster can at most block 16 clusters (see Figure 3(b)). The maximum interference happens when a user in the corner of a cluster transmits a file to a user in the opposite corner. So, we have $E[L]\geq\frac{E[G]}{17}$ where $E[G]$ is the expected number of good clusters. Since we want to find the lower bound for $E[L]$, we can limit users to communicate with users in virtual clusters they belong to. Therefore, we have

$$E[G]\geq\frac{2}{r(n)^2}\sum_{k=0}^{n}\Pr[\text{good}|k]\Pr[K=k],$$

where $\frac{2}{r(n)^2}$ is the total number of virtual clusters. K is the number of users in the cluster, which is a binomial random variable with n trials and probability of $\frac{r(n)^2}{2}$, *i.e.,* $K=B(n,\frac{r(n)^2}{2})$. $\Pr[K=k]$ is the probability that there are k users in the cluster and $\Pr[\text{good}|k]$ is the probability that the cluster is good conditioned on k. The probability that a cluster is good depends on what users cache. Therefore,

$$E[G]\geq\frac{2}{r(n)^2}\sum_{k=0}^{n}\Pr[K=k]\times\sum_{\{\mathbf{u}\,|\,|\mathbf{u}|=k\}}\Pr[\text{good}|\mathbf{u},k]\Pr[\mathbf{U}=\mathbf{u}],\qquad(2)$$

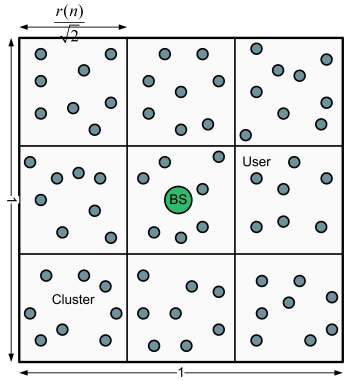
where $\mathbf{U}$ is a random vector of stored files by users in the cluster. $\mathbf{u}$ is a realization of $\mathbf{U}$ and $|\mathbf{u}|$ denotes the length of vector $\mathbf{u}$. The ith element of $\mathbf{u}$ denoted by $\mathbf{u_i}\in\{1,2,3,\ldots,m\}$ indicates what user i in the cluster stores.

For each $\mathbf{u}$, we define a value:

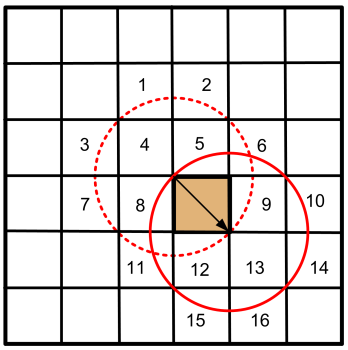
$$v(\mathbf{u})=\sum_{\mathbf{i}\in\tilde{\mathbf{u}}}\mathbf{f_i},$$

where $\tilde{\mathbf{u}}=\cup_{j=1}^{|\mathbf{u}|}\mathbf{u_j}$ and $\cup$ is the union operation. Actually $v(\mathbf{u})$ is the sum of popularities of the union of files in $\mathbf{u}$. The cluster is considered to be good if at least a user i in the cluster requests one of the files in $\tilde{\mathbf{u}}-\{\mathbf{u_i}\}$. Note the possibility of *self-requests*, *i.e.,* a user might find the file it requests in its own cache; in this case clearly no D2D communication will be activated by this user. Accounting for these self-requests, the probability that user i finds its request files locally within the cluster is $(v(\mathbf{u})-f_{\mathbf{u_i}})$. Thus, we obtain:

$$\Pr[\text{good}|\mathbf{u},k]\geq 1-(1-(v(\mathbf{u})-\max_i f_{\mathbf{u_i}}))^k.\qquad(3)$$



(a)



(b)

Fig. 3. a) Dividing cell into virtual clusters. b) In the worst case, a good cluster can block at most 16 clusters. In the dashed circle, receiving is not possible and in the solid circle, transmission is not allowed.

Let us only consider cases where at least one user in the cluster caches file 1 (the most popular file). Then, from (2) and (3), the following lower bound is achieved:

$$E[G] \geq \frac{2}{r(n)^2} \sum_{k=1}^{n} \Pr[K=k] \times \sum_{\mathbf{u}\in\mathbf{x}} 1-(1-(v(\mathbf{u})-f_1))^k \Pr[\mathbf{U}=\mathbf{u}]. \quad (4)$$

where $\mathbf{x} = \{\mathbf{u} \,|\, |\mathbf{u}| = k \text{ and } 1 \in \mathbf{u}\}$. Let us further define a random variable V which is sum of popularities of the union of files stored by users in the cluster. Then, in equation (4), we can take the expectation with respect to V, *i.e.,*

$$E[G] \geq \frac{2}{r(n)^2} \sum_{k=1}^{n} \Pr[K=k] E_V[1-(1-(V-f_1))^k | A_1^k] \geq \frac{2}{r(n)^2} \sum_{k=1}^{n} \Pr[K=k] E_V[(V-f_1)|A_1^k],$$

where A_1^k is the event that at least one of k users in the cluster caches file 1 and $E_V[.]$ is the expectation with respect to V. Let $A_{1,h}^k$ for $1 \leq h \leq k$ denote the event that h users out of k users in the cluster cache file 1. Then, we get:

$$E[G] \geq \frac{2}{r(n)^2} \sum_{k=1}^{n} \Pr[K=k] \sum_{h=1}^{k} E_V[(V-f_1)|A_{1,h}^k] \times \binom{k}{h} (p_1)^h (1-p_1)^{k-h}, \quad (5)$$

where p_j represents the probability that file j is cached by a user based on Zipf distribution with exponent γ_c. To calculate $E_V[(V-f_1)|A_{1,h}^k]$, we define an indicator function $\mathbf{1}_j$ for each file $j \geq 2$. $\mathbf{1}_j$ is equal to 1 if at least one user in the cluster stores file j. Hence,

$$E_V[(V-f_1)|A_{1,h}^k] = E[\sum_{j=2}^{m} f_j \mathbf{1}_j | A_{1,h}^k] = \sum_{j=2}^{m} f_j (1-(1-p_j)^{k-h}).$$

Substituting $E_V[(V-f_1)|A_{1,h}^k]$ in (5) and limiting the interval of k, we can obtain:

$$E[G] \geq \frac{2}{r(n)^2} \sum_{k\in I} \Pr[K=k] \times \sum_{h=1}^{k}\sum_{j=2}^{m} f_j (1-(1-p_j)^{k-h}) \binom{k}{h} (p_1)^h (1-p_1)^{k-h}, \quad (6)$$

where $0<\delta<1$ and $I = [nr(n)^2(1-\delta)/2, nr(n)^2(1+\delta)/2]$. Define $k^* \in I$ such that it minimizes the expression in the last line of (6). Considering that $r(n) = \Theta(\sqrt{\frac{1}{n}})$, k^* is $\Theta(1)$. Then from (6), we have:

$$E[G] \geq \frac{2}{r(n)^2} \Pr[k \in I] \sum_{h=1}^{k^*}\sum_{j=2}^{m} f_j (1-(1-p_j)^{k^*-h}) \times \binom{k^*}{h} (p_1)^h (1-p_1)^{k^*-h} \quad (7)$$

$$\geq \frac{2}{r(n)^2} (1-2e^{-nr(n)^2\delta^2/6}) \sum_{h=k^*p_1(1-\delta_1)}^{k^*p_1(1+\delta_1)} \Big[\binom{k^*}{h} \times \sum_{j=2}^{m} f_j (1-(1-p_j)^{k^*-h}) (p_1)^h (1-p_1)^{k^*-h} \Big], \quad (8)$$

where $0<\delta_1<1$. We apply the Chernoff bound in (7) to derive (8) [16]. Since the exponent $nr(n)^2\delta^2/6$ is $\Theta(1)$, we can select the constant c_1 such that the term $(1-2e^{-nr(n)^2\delta^2/6})$ becomes positive.

Let us define $h^* \in [k^*p_1(1-\delta_1), k^*p_1(1+\delta_1)]$ such that it minimizes the expression in the last line of (8). From (1) and lemma 1, p_1 is $\Theta(1)$ and as a result, h^* is also $\Theta(1)$. Using the Chernoff bound in (8), we get:

$$E[G] \geq \frac{2}{r(n)^2} (1-2e^{-nr(n)^2\delta^2/6})(1-2e^{-k^*p_1\delta_1^2/3}) \times \binom{k^*}{h^*} (p_1)^{h^*} (1-p_1)^{k^*-h^*} \sum_{j=2}^{m} f_j (1-(1-p_j)^{k^*-h^*}). \quad (9)$$

$k^* - h^*$ should be greater than 1 which results in a constant lower bound for c_1. The second exponent, *i.e.,* $k^* p_1 \delta_1^2/3$ is $\Theta(1)$. The term $(1 - 2e^{-k^* p_1 \delta_1^2/3})$ is a positive constant if $c_1 \geq \frac{3 \ln 2\zeta(\gamma_c)}{\delta_1^2(1-\delta)}$, where $\zeta(\gamma) = \sum_{j=1} \frac{1}{j^\gamma}$ is the Riemann zeta function [17]. Further, the summation in (9) satisfies

$$\sum_{j=2}^{m} f_j (1 - (1 - p_j)^{k^* - h^*}) > \sum_{j=2}^{m} f_j p_j.$$

To show that $E[G]$ scales linearly with n, the term $\sum_{j=2}^{m} f_j p_j$ should not be vanishing as n goes to infinity. It can been shown that if $\gamma_r, \gamma_c > 1$, $\sum_{j=2}^{m} f_j p_j = \Theta(1)$ (see lemma 1).

Lemma 1: If $\gamma > 1$, $a = o(b)$, and $a = \Theta(1)$, then $H(\gamma, a, b) = \Theta(1)$ and $\sum_{j=a}^{b} f_j p_j = \Theta(1)$ where $H(\gamma, a, b) = \sum_{j=a}^{b} \frac{1}{i^\gamma}$.

The proof is omitted due to lack of space.

REFERENCES

[1] "http://www.cisco.com/en/us/solutions/collateral/ns341/ns525/ns537 /ns705/ns827/white_paper_c11-520862.html."

[2] N. Golrezaei, K. Shanmugam, A. Dimakis, A. Molisch, and G. Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," in *INFOCOM*. IEEE, 2012.

[3] ——, "Wireless video content delivery through coded distributed caching," in *ICC*. IEEE, 2012.

[4] P. Gupta and P. Kumar, "The capacity of wireless networks," *Information Theory, IEEE Transactions on*, vol. 46, no. 2, pp. 388–404, 2000.

[5] M. Penrose and O. U. Press, *Random geometric graphs*. Oxford University Press Oxford, 2003, vol. 5.

[6] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, 2007, pp. 1–14.

[7] "http://traces.cs.umass.edu/index.php/network/network."

[8] N. Golrezaei, A. Dimakis, and A. Molisch, "Asymptotic throughput of base station assisted device-to-device communications," pp. 382–390, to be submitted for publication.

[9] A. Molisch, *Wireless communications*. Wiley, 2011.

[10] E. Lawler, J. Lenstra, A. Kan, and E. U. E. Institute, "Generating all maximal independent sets: Np-hardness and polynomial-time algorithms," *SIAM J. Comput.*, vol. 9, no. 3, pp. 558–565, 1980.

[11] A. Ozgur, O. Lévêque, and D. Tse, "Hierarchical cooperation achieves linear capacity scaling in ad hoc networks," in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*. IEEE, 2007, pp. 382–390.

[12] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," in *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3. IEEE, 2001, pp. 1360–1369.

[13] M. Franceschetti, M. Migliore, and P. Minero, "The capacity of wireless networks: information-theoretic and physical limits," *Information Theory, IEEE Transactions on*, vol. 55, no. 8, pp. 3413–3424, 2009.

[14] Y. Chen, C. Caramanis, and S. Shakkottai, "On file sharing over a wireless social network," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 249–253.

[15] N. Golrezaei, A. Molisch, and A. Dimakis, "Base station assisted device-to-device communications for high-throughput wireless video networks," *submitted for publication*.

[16] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *The Annals of Mathematical Statistics*, vol. 23, no. 4, pp. 493–507, 1952.

[17] J. Conrey, "The riemann hypothesis," *Notices of the AMS*, vol. 50, no. 3, pp. 341–353, 2003.